不确定性环境下的矛盾知识主张识别研究

■ 郭进京 黄奇
南京大学信息管理学院 南京 210023

摘 要：[目的/意义]科学研究(尤其是医学研究)充满了不确定性,识别研究中的矛盾知识主张将有助于识别“科学分歧/不一致的科学结论”,推动潜在变革性科学研究的识别和相关研究的完善。[方法/过程]以阿尔茨海默病为例,将 PubMed 文摘数据作为数据源,采用 SemRep 工具进行三元组抽取。制定表征矛盾意义的知识主张识别规则,依据不确定性程度高低对来源语句进行划分,分别采用单句识别和跨语句识别两种途径,识别以三元组形式表示的、具有矛盾意义的医学研究知识主张。[结果/结论]从来源于 PubMed 的 6 574 篇医学文摘中共计识别出 49 组(涉及 277 对三元组)矛盾知识主张。阿尔茨海默病在诊断和治疗方面的研究仍存在部分争议和矛盾,有待进一步验证。矛盾知识主张识别为潜在变革性医学研究前沿发现提供新思路,可用于基于知识计量的知识发现和为知识图谱可信度计算提供参考。

关键词：知识发现 不确定性 矛盾知识 SPO 阿尔茨海默病
分类号：G250
DOI：10.13266/j.issn.0252-3116.2021.20.013

1 引言

“不确定性”(uncertainty)一词最早出现在由英国著名政治经济学家詹姆斯·穆勒于 1835 年发表的 *Law Reform*[1]一文中,指出法律条文的错误表述会带来歧义和不确定性。在量子力学中,A-B 效应(阿哈罗诺夫-玻姆效应)意味着微观世界本身具有不确定性[2]。微观客体的内在不确定性必然决定描述量子客体的知识具有不确定性。知识是对客体或客观过程的反映,因此描述客观世界运动规律的科学知识具有不确定性。医学研究所依赖的所有临床数据(包括诊断结果、患者提供的病史和体检结果等)往往是不完整的,都具有不确定性。有研究表明,生物医学领域常用数据库 PubMed 中的论文摘要中,约 11% 的句子使用了模糊修饰词或推测性表达[3],具有高度的不确定性。Cell 杂志在 2021 年的首刊发刊词中指出:人们对于科学进程的理解不是直线式发展的,每一个科学结论都是用不确定性衡量的,更好的实验否定早期的假设不是科学的失败,而是一种成功;一个好的科学家应该正确面对与其之前的结论或假设相矛盾的更好证据[4]。识别医学研究中的“科学分歧/不一致的科学结论”有助于发现研究的争议或矛盾之处,并据此寻找更好的证据,推动医学研究在争论中向前发展。

阿尔茨海默病(Alzheimer disease,AD)是痴呆(dementia)的一种,根据阿尔茨海默病国际协会(Alzheimer's Disease International, ADI)网站数据[5],全球目前已有 5 000 余万人患有阿尔茨海默病或其他类型痴呆症,预计到 2050 年这个数字将达到 1.52 亿。根据中国老年医学会网站 2019 年数据[6],我国有 1 000 万左右的阿尔茨海默病患者,是全球阿尔茨海默病(AD)患者数量最多的国家,预计到 2050 年患病人数将超过 4 000 万。笔者基于不确定性原理和前人关于矛盾/冲突谓词选取的研究基础,制定表征矛盾/冲突意义的三元组的识别规则,以阿尔茨海默病为例,从海量医学文本中抽取出以三元组形式表示的、具有矛盾争议的诊断和治疗类临床知识主张,以期识别出该领域内可能具有突破性研究倾向的知识点,为医学领域研究前沿的识别提供新的方式,挖掘潜在的新的诊断和治疗方法。

2 矛盾/冲突知识主张识别的方法与工具

2.1 理论基础及相关研究进展

在 *The structure of scientific revolutions: 50th anniver-*

作者简介：郭进京(ORCID:0000-0003-1590-8014),博士研究生;黄奇(ORCID:0000-0003-2394-148X),教授,博士,博士生导师,通讯作者,E-mail:huangq@nju.edu.cn。
收稿日期:2021-05-17 修回日期:2021-08-09 本文起止页码:123-134 本文责任编辑:徐健

sary edition[7]一书中，托马斯·库恩根据范式转移程度，将科学研究分为常规科学研究和变革性研究：常规科学研究旨在“解谜”，致力于求解当前的某一知识领域的谜题，趋于发现那些它期待发现的事情，是一个不断提出假设推测并进行验证的过程；变革性研究旨在通过不断地发现新现象和反常现象（即“矛盾/冲突”），并尝试解释或解决这类反常，最终产生出新的理论。这两类研究的范式转移程度与不确定性程度的对应关系如图 1 所示[8]。常规科学研究对应着验证假设推测的过程，变革性研究对应着解决争议矛盾的过程。根据牛津词典（第二版）的定义，矛盾（contradiction）是指存在两个或两个以上不相容主张来描述同一主张[9]。C. M. Chen 和 M. Song 认为不确定性科学知识主要是指处于假设推测或争议矛盾状态的知识，科学的进步必须处理有争议的、矛盾的、不完整的和不一致的信息[10]。不确定性原理为识别科学研究过程中的矛盾/冲突提供了理论基础，进而为识别潜在的变革性科学研究提供了可能的新思路。

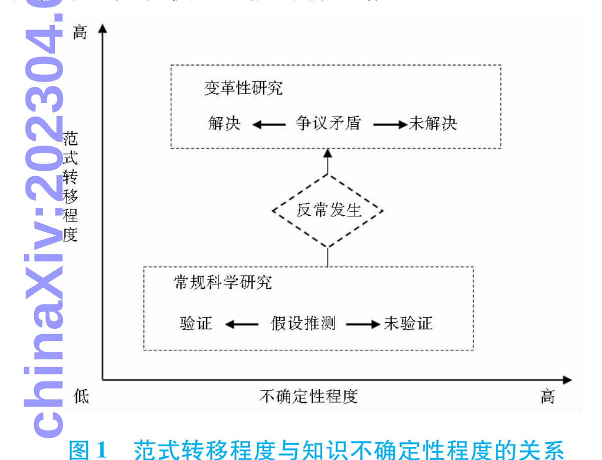


图 1 范式转移程度与知识不确定性程度的关系

关于知识主张（knowledge claims），M. W. McElroy 在《新知识管理：复杂性、学习和可持续创新》一书中将其定义为“关于何种潜在行为可能导致预期结果的猜想、断言、论证或理论，旨在缩小当前状态和目标状态间的差距”[11]。K. Peters 等[12]指出知识主张的类型多样，包括理论、模型、声明、方法和设计。一方面，知识主张涉及如何执行任务或改变事物；另一方面，知识主张可能基于其他知识主张，是关于事物现在和将来发挥作用或运行的方式。科学研究中的知识主张是指研究人员想要呈现给读者的最重要的内容，可以从研究发现或研究结果中理解出的总体结论或结果，具有相反的断言值（assertion values）的知识主张被认为是潜在的矛盾知识主张[13]。综合上述定义，本研究中的“矛盾知识主张”是指从科学研究的结论或结果中获取到的具有相反蕴涵的猜想、理论、模型、声明、方法和设计。

矛盾/冲突的产生可分为语法层面的冲突和语义层面的冲突，语法层面的冲突主要由各类语言结构导致，如反义词、否定（包括直接否定和间接否定）、数字或时间不匹配、对事实真实程度的隐含假定（包括使役动词、隐含动词和非使役动词）、情态表达（包括可能性情态动词和必要性情态动词）、结构（包括词汇结构和句法结构）、背景知识等[14]；语义层面的矛盾冲突主要来源于释义和文本蕴涵的差异[15]。相应的，矛盾/冲突识别的方法可以分为两类：对于语法层面的矛盾冲突，需要对涉及到的否定、对比和反义词等语言特征进行分析；对于语义层面的矛盾冲突，则可以在去除否定命题后衡量文本间的文本蕴涵是否成立，如果成立则说明该文本对是冲突的[16]。“科学分歧/不一致的科学结论”是指两个知识主张涉及同一事物但又不可能同时为真时，被认为是互相矛盾的。围绕着“科学分歧/不一致的科学结论”这一类型不确定知识的测度，部分学者从引用语句切入，分析科研人员对科学主张和论断的不确定性程度评价。H. Kilicoglu 等[17]根据引用情感与被引临床研究结论的一致程度，对临床医学研究文献的引用语句进行情感分析，负向情感表示与被引临床研究结论不一致；D. Murray 等[18]构建了表征科学分歧的线索词（contradict、conflict）和过滤词（studies、results），根据线索词和过滤词测度引用语句中的科学分歧。基于引用语句的分析，是基于概念共现的相关关系的探测，无法识别出结构化的、表示因果关系的不一致结论。此外，也有学者从文本内容分析和知识单元（knowledge unit）角度切入开展矛盾性医学知识识别：对于语句内识别，依据语句中是否包含表示矛盾/冲突的提示词（如 controversial、contradictory、conflicting 等）来识别语句内的带有矛盾/冲突表征的三元组（主语－谓词－宾语，是细粒度表征知识对象及其关系的知识单元）；对于跨语句识别，主要是基于三元组中“主宾相同、谓语相反”规则识别矛盾性医学知识，不同的是研究人员对谓词的选取和分类。A. Alamri[13]将谓词分为主动/导致类（如 AUGMENTS、CAUSES）、被动/抑制类（如 DISRUPTS、PREVENTS）和其他类（如 ADMINISTERED_TO、OCCURS_IN）；G. Rosemblat 等[19]将谓词分为有因果意义的 4 对相反谓词对和无因果意义的 4 对相反谓词对；J. M. G. Pinto 等[20]则选取了 7 种谓词来识别矛盾性知识和多样性知识；X. Y. Li 等[21]将谓词分为激励类（excitatory）和抑制类

(inhibitory)并设定了识别规则来发现矛盾/冲突类知识主张(knowledge claims)和多样性知识主张。列出表征矛盾/冲突的谓词对的方式,可以清晰地理解识别出的三元组的因果关系,但识别出的知识主张数量会受限于识别规则的数量;将谓词划分为激励类和抑制类两个相互矛盾/冲突的群组的方式,可以识别出更多的矛盾/冲突知识主张,但不能清晰地获取三元组中的因果关系。

综合上述研究现状,笔者认为不同的知识主张其所表达的知识的不确定性程度是不一样的,需先依据源语句中是否包含表征不确定性程度的提示词进行不确定性程度的区分,进而再以从源语句中抽取出的三元组为判别对象,通过抽取出的谓词确定实体之间的关系,依据规则判断两对三元组是否是矛盾/冲突的,将分析角度从相关关系转向因果关系,更便于解释矛盾/冲突的由来。

2.2 三元组抽取工具

常用的三元组抽取工具有开放信息抽取工具(Open Information Extraction, OpenIE)[22]、Goldfire[23]、Semrep[24]等,分别面向开放信息、专利和生物医学领域。笔者采用 SemRep 工具从生物医学文本中抽取语义关系,SemRep 抽取的三元组主语和宾语都是 UMLS 超级叙词表(metathesaurus)中带有特定语义类型的概念,谓词来源于扩展后的 UMLS 语义网络关系类型。SemRep 既可以从语法层面对涉及到的否定、对比和反义词等语言特征进行分析,也能从语义层面依据识别出的语义关系来揭示文本蕴涵。SemRep 工具能够识别出的语义关系类型主要有:临床相关(如 treats、diagnoses、process_of),分子相互作用(如 interacts_with、inhibits、stimulates),疾病病因学(如 associated_with、causes、predisposes),药物基因组学(如 affects、augments、disrupts),静态关系(isa、part_of、location_of)等,能够满足本研究对语义谓词类型的需要。SemRep 针对不同类型的语义关系的识别精确度较高,针对特定谓词领域(如药物基因组学)和特定语言结构(如比较谓词)的准确率分别达到了 76%[25]和 96%[26],在基于标准的基准语料库 CDR(chemical-disease relation)的句子层级的评估中,将谓词限定在"causes、affects、augments、stimulates、predisposes 和 associated_with"等带有因果关系的语义谓词,抽取准确率达到 90%[27],能够满足本研究对抽取出的三元组的准确性的需要。SemRep 提供了网页交互、批量处理、API 远程调用和本地安装 4 种使用方式,笔者采用批量处理方式进行三元组抽取,所用工具版本为 1.8 版本,输出结果为 txt 格式。

2.3 矛盾/冲突知识主张识别方法与规则

矛盾/冲突知识主张识别的总体识别方法路径如图 2 所示。在进行单句和跨语句识别前,需要对源语句的不确定性程度进行判定,如果源语句中含有"may""could""might"等表示高不确定性程度的提示词,则将其划入到不确定性程度较高的知识主张识别过程中,反之则划入不确定性程度较低的知识主张识别过程。

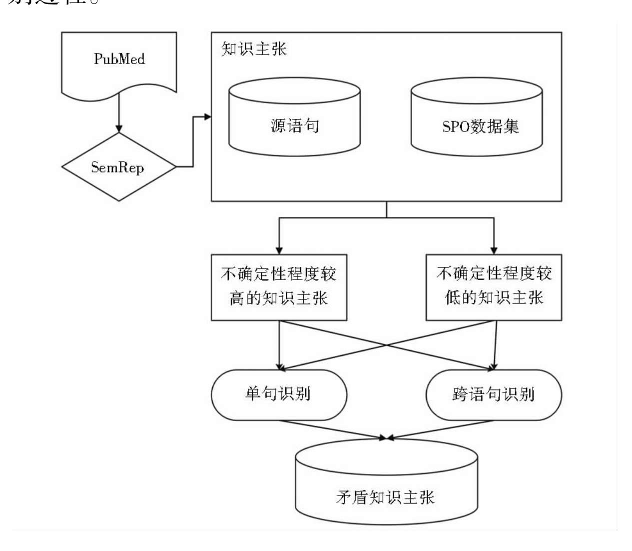


图 2 矛盾知识主张识别方法路径

矛盾/冲突知识主张识别主要分为单句识别和跨语句识别两类[21]。

对于单句识别,可以通过查找句子中是否包含"controversial""debatable""contentious""contradictory""contrary""opposing"和"conflicting"等明显的提示词来判定该句子是否为蕴涵矛盾/冲突的语句,然后提取句子中的三元组。

笔者将三元组的表示形式定义为 $T_in = (C_i1, R_in, C_i2)$,从句子 S_i 中抽取获得,C_i1 为三元组 T_in 的主语,C_i2 为三元组 T_in 的宾语,R_in 为三元组 T_in 的谓词。定义提示词集合 $M = \{W_1, W_2, \cdots, W_n\}$,N 为构成 S_i 的所有单词的集合 $\{N_i1, N_i2, \cdots, N_in\}$,定义 U 为蕴涵矛盾/冲突的三元组集合。单句识别的规则如下:

Input S_i //输入语句 S_i

If $N \cap M \neq \varphi$ //如果 S_i 中含有集合 M 中的任一提示词

Then $T_in = (C_i1, R_in, C_i2) \in U$ //则从 S_i 中抽

取出的三元组 T_in 为蕴涵矛盾/冲突的三元组

对于跨语句识别，可以根据以下规则进行判定：从科学主张中抽取两个及以上的"主语－谓词－宾语(Subject-Predicate-Object, SPO)"三元组，这些三元组有着共同的主语和宾语，但谓词的语义类型相反。定义另一对三元组 $T_jm = (C_j1, R_jm, C_j2)$，从句子 S_j 中抽取获得，定义矛盾语义谓词对集合 $P = \{(R_i1, R_j1), (R_i2, R_j2)\cdots, (R_in, R_jm)\}$，定义 V 为互相矛盾/冲突的三元组对集合，跨语句识别规则如下：

Input $T_i n$, T_jm //输入三元组 $T_i n$ 和 T_jm

If $C_i1 = C_j1$, $C_i2 = C_j2$, $(R_in, R_jm) \in P$ //如果三元组 T_i 和 T_j 中的主语和宾语相同，且谓词对来自于集合 P 中的任一元素

Then$(T_in, T_jm) \in V$ //则分别从 S_i 和 S_j 中抽取出的三元组 T_in 和 T_jm 互为矛盾/冲突三元组对

根据需要，笔者从 SemRep 工具能够识别出的众多类型的语义关系中选取了与临床诊断和治疗密切相关的语义关系类型，主要谓词包括：treats、causes、prevents、predisposes、augments、disrupts、stimulates、inhibits，这些谓词的否定形式(如 neg_treats)也被考虑在内，最终形成的相反谓词对如表 1 所示，包括涉及因果关系的谓词对和缺少因果关系的谓词对。如果两对三元组含有相同的主语和宾语，且谓词为表 1 中所列出的表示矛盾关系的谓词对，则判定该两对三元组为矛盾知识主张对。

表 1　研究所采用的表示矛盾关系语义谓词对

涉及因果关系的谓词对			缺少因果关系的谓词对		
treats	vs	causes	treats	vs	neg_treats
treats	vs	predisposes	causes	vs	neg_causes
prevents	vs	causes	predisposes	vs	neg_predisposes
prevents	vs	predisposes	prevents	vs	neg_prevents
augments	vs	disrupts	augments	vs	neg_augments
stimulates	vs	inhibits	stimulates	vs	neg_stimulates
			disrupts	vs	neg_disrupts
			inhibits	vs	neg_inhibits

具体而言，对于单句识别，提示词集合 M = {controversial, debatable, contentious, contradictory, contrary, opposing, conflicting}。对于跨语句识别，矛盾语义谓词对集合 P = {(treats, causes), (treats, predisposes), (prevents, causes), (prevents, predisposes), (augments, disrupts), (stimulates, inhibits), (treats, neg_treats), (causes, neg_causes), (predisposes, neg_predisposes), (prevents, neg_prevents), (augments, neg_augments), (stimulates, neg_stimulates), (disrupts, neg_disrupts), (inhibits, neg_inhibits)}。

利用上述规则识别出的一对矛盾知识主张例句如表 2 所示，其中语句 31895707. ab. 1[28] 和语句 32476013. ti. 1[29] 因相反的谓词对"predisposes"和"neg_predisposes"而表达相矛盾的知识主张，即"质子泵抑制剂(proton pump inhibitors)"会/不会增加"痴呆(dementia)"的风险。

表 2　一对矛盾的知识主张示例

源语句	主语	谓词	宾语
31895707. ab. 1 Introduction: Long-term use of *proton pump inhibitors* (*PPIs*) has been associated with a wide variety of potentially serious adverse effects including a possible *increased risk of dementia*.	proton pump inhibitors	predisposes	dementia
32476013. ti. 1 *Proton pump inhibitors do not increase* the *risk* of *dementia*: a systematic review and meta-analysis of prospective studies.		neg_predisposes	

3　矛盾/冲突知识主张识别结果

3.1　数据获取与处理

科学文献是科学知识的主要载体，可理解为针对科学问题、提供证据支持的知识主张记录。PubMed (https://pubmed.ncbi.nlm.nih.gov/)是一个提供生物医学方面的科学文献引文及摘要信息的数据库，由美国国立医学图书馆(National Library of Medicine, NLM)开发，目前包含了来自 MEDLINE、生命科学期刊和在线书籍的 3 200 多万条生物医学文献引文和摘要信息，是医学领域科研人员频繁使用的文献检索平台。因此，笔者将 PubMed 作为矛盾知识主张识别的数据来源。对于不同类型的科技文献，其所提供的证据/主张的可信度是不同的[19]。在现实世界中，与案例研究相比，随机对照实验(Randomized Controlled Trial, RCT)和系统综述类的研究具有更高的临床证据水平[30]，能够从支撑语句中抽取出具有更高可信度和准确度的矛盾/冲突发现。因此，笔者以"Alzheimer's Disease"为主题，将文献类型限定为具有较高可信度和准确度的临床指南、元分析文章、临床试验类文章、对比分析文章等，检索近 10 年内发表的英文文章，检索时间为 2021 年 1 月 26 日，扩展后的数据检索式为：

("alzheimer disease"[MeSH Terms] OR ("alzheimer"[All Fields] AND "disease"[All Fields]) OR "alzheimer disease"[All Fields] OR ("alzheimer s"[All Fields] AND "disease"[All Fields]) OR "alzheimer s disease"[All Fields]) AND ((y_10[Filter]) AND (clinicalstudy[Filter] OR clinicaltrial[Filter] OR clinicaltrialphasei[Filter] OR clinicaltrialphaseii[Filter] OR clinicaltrialphaseiii[Filter] OR clinicaltrialphaseiv[Filter] OR comparativestudy[Filter] OR controlledclinicaltrial[Filter] OR guideline[Filter] OR meta-analysis[Filter] OR multicenterstudy[Filter] OR practiceguideline[Filter] OR pragmaticclinicaltrial[Filter] OR randomizedcontrolledtrial[Filter] OR systematicreview[Filter]) AND (english[Filter]))

根据上述检索式,笔者从 PubMed 中共检索到近 10 年发表的 6 574 篇与"Alzheimer's Disease"主题相关的文章,将文章摘要信息(含标题、PMID 号)导入 SemRep 工具,经过分词、词汇/句法分析、短语映射、语义谓词归一化、概念语义类型约束等步骤后,共计得到 45 674 对三元组,涉及 76 635 条句子。根据需要,笔者依据句子中是否含有"may""could""might"等表示高不确定性程度的提示词[31],将 76 635 条句子划分为两类:一类为不确定性程度较高的句子(含有"may""could""might"等提示词),另一类为不确定性程度较低的句子(不含"may""could""might"等提示词),分别按照 2.3 中的规则进行矛盾/冲突知识主张的单句识别和跨语句识别,具体实验流程及结果如图 3 所示。在去除不能从中抽取出三元组的源语句之后,最终纳入分析的源语句及三元组数量如表 3 所示。在根据规则识别出矛盾/冲突三元组对之后,由专家对抽取结果进行判读并得到最终结果。

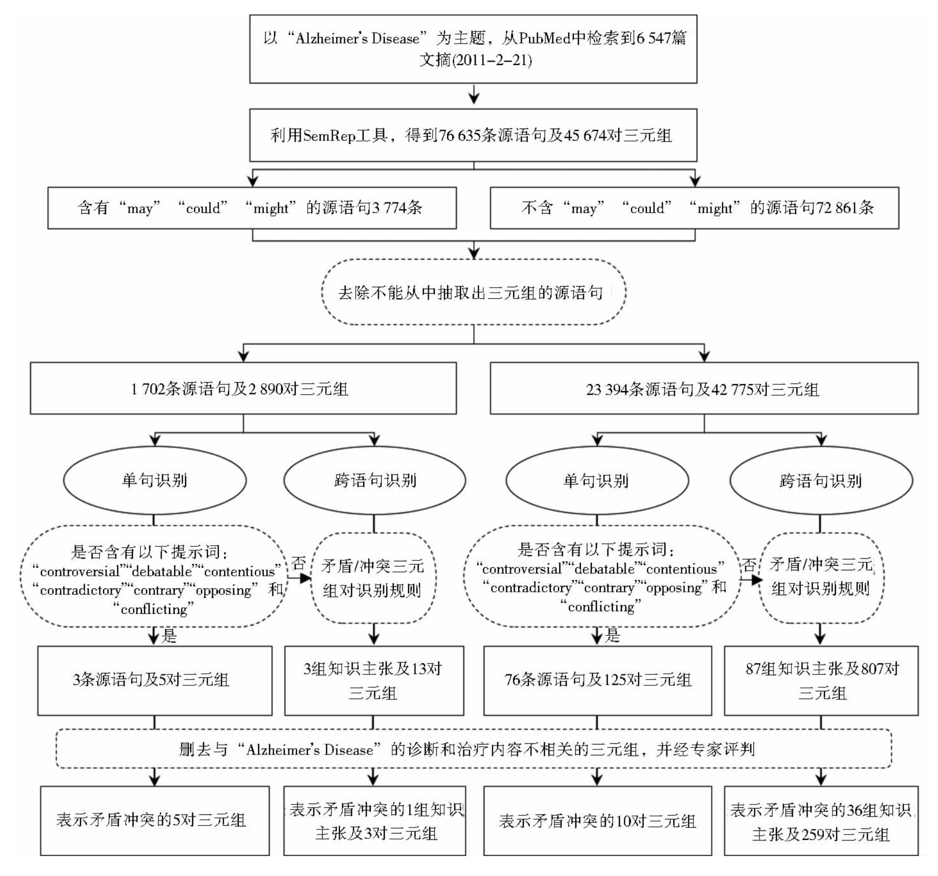


图 3　矛盾/冲突主张识别实验流程及结果

表 3　实验数据的基本描述性统计　(单位/条)

数据类型	源语句数量	三元组数量
初始数据	76 635	45 674
不确定性较高的知识主张	1 702	2 890
不确定性较低的知识主张	25 394	42 775

3.2　不确定性程度较高的矛盾/冲突知识主张识别

根据句子中是否含有"may""could""might"等提示词,笔者从 76 635 条句子中筛选出 3 774 条符合条件的句子,得到来源于 1 702 条有效语句(有的语句虽含有表示高不确定性的词语,但并不能从中抽取出三

元组)的2 890对三元组。以下将从单句识别和跨语句识别分别进行分析。

3.2.1 不确定性程度较高的矛盾/冲突知识主张单句识别

根据2.3中所设定的识别规则,首先查找1 702条句子中是否包含以下提示词:"controversial""debatable""contentious""contradictory""contrary""opposing"和"conflicting",根据这些提示词来判定该句子是否为蕴涵矛盾/冲突的语句,然后再提取句子中的三元组。经过筛选和专家判断,共计得到来源于3条语句的5对三元组(见表4)。例如,从语句30098211. ab. 1[32]中可以看出,明苯二氮卓类药物(benzodiazepines)与痴呆(dementia)的关系是存疑的,需进一步验证。

表4 单句内不确定性程度较高的矛盾/冲突知识主张识别结果

出版时间	源语句	三元组	不确定词	矛盾/冲突词
2016.11	27845333. ab. 9 However, ***contrary*** to observational findings, coffee ***may*** not have beneficial effects on T2DM, depression or Alzheimer's disease.	coffee-***neg_treats***-Alzheimer's Disease; coffee-***neg_treats***-depressive disorder; coffee-***neg_treats***-diabetes mellitus, non-insulin-dependent	***may***	***contrary***
2017.09	28349979. ab. 1 Background: Androgen deprivation therapy (ADT) to treat prostate cancer ***may*** be associated with an increased risk of dementia, but existing studies have shown ***conflicting*** results.	androgen therapy-***treats***-malignant neoplasm of prostate	***may***	***conflicting***
2018.10	30098211. ab. 1 Study objective: The use of benzodiazepines and the development of dementia is ***controversial***, with studies indicating that benzodiazepines could be either a protective factor or a risk factor for dementia, or no association ***may*** exist be.	benzodiazepines-***predisposes***-dementia	***may***	***controversial***

3.2.2 不确定性程度较高的矛盾/冲突知识主张跨语句识别

在去除3.2.1中识别出的5对三元组后,根据2.3中所设定的规则——"主语、宾语相同,谓词语义类型相反"对剩余的2 885对三元组进行跨语句识别,共计识别出3组表示矛盾/冲突知识主张的三元组,涉及13对三元组,经过专家评判,最终保留一组(见表5)。语句21647938. ab. 3[33]指出,脑源性神经营养因子(Brain-Derived Neurotrophic Factor, BDNF)有可能(may)会促进AD的发展(Brain-Derived Neurotrophic Factor-predisposes-Alzheimer's Disease),而31518516. ab. 8[34]则指出BNDF可能不适合(may not)作为诊断AD的最佳生物标记物,即与AD的关系不显著(Brain-Derived Neurotrophic Factor-neg_predisposes-Alzheimer's Disease),这两对三元组的知识主张是存在争议的。

表5 跨语句不确定性程度较高的矛盾/冲突知识主张识别结果

出版时间	源语句	矛盾/冲突三元组
2011.06	21647938. ab. 3 Dysregulation of brain-derived neurotrophic factor (BDNF) has been implicated in deficits of long-term potentiation and cognition and ***may*** contribute to the development of Alzheimer's disease (AD).	Brain-Derived Neurotrophic Factor-***predisposes***-Alzheimer's Disease
2019.09	31518516. ab. 8 However, the ROC curve analysis revealed that the peripheral BDNF levels ***may not*** be an optimal biomarker potentially for AD and MCI diagnosis with a lower AUC (AD: 0.707; MCI: 0.573), less sensitivity (AD: 66.67%; MCI: 50.00%) and poor specificity (AD: 93.33%; MCI: 83.33%).	Brain-Derived Neurotrophic Factor-***neg_predisposes***-Alzheimer's Disease
2019.09	31518516. ab. 9 These results suggested that AD or MCI is accompanied by reduction of peripheral BDNF, but the levels of circulating BDNF ***may*** *not be suitable as a diagnostic marker for AD and MCI.*	*Brain-Derived Neurotrophic Factor-**neg_predisposes**-Alzheimer's Disease*

3.3 不确定性程度较低的矛盾/冲突知识主张识别

除去3.2中不确定性程度较高的3 774条句子,不含有"may""could""might"等提示词的语句共有72 861条。经过整理,去除不能从中识别出三元组的语句,共计得到来源于25 394条有效语句的42 775对三元组。以下将从单句识别和跨语句识别分别进行分析。

3.3.1 不确定性程度较低的矛盾/冲突知识主张单句识别

采用3.2.1中的筛选方式,首先查找25 394条句子中是否包含以下提示词:"controversial""debatable""contentious""contradictory""contrary""opposing"和"conflicting",根据这些提示词来判定该句子是否为表示矛盾/冲突的语句,然后再提取句子中的三元组。经

过初次筛选,得到 76 条语句含有上述提示词,从中抽取出 125 对三元组。经过二次筛选,共计得到与 AD 的诊断和治疗相关的来源于 12 条语句的 13 对三元组。经过专家判读,最终确定来源于 9 条语句的 10 对矛盾/冲突三元组(见表 6)。例如,现有的研究指出,抗氧化剂(antioxidants)在 AD 的预防方面的作用,血管内皮生长因子(vascular endothelial growth factor, VEGF)基因是否是 AD 的诱发因素,补充剂(supplementation)是否可以预防 AD,都是存在争议的。组织蛋白酶 D 基因(Cathepsin D, CTSD)、铝暴露(aluminum)、羟基甲基戊二酰辅酶 A(Hydroxymethylglutaryl-CoA, HMG-CoA)还原酶抑制剂(reductase inhibitors)、乙醇(ethanol)等对 AD 诱发作用的研究是存在矛盾的,银杏叶提取物 EGb761、美金刚(memantine)对 AD 的治疗作用的研究也是存在矛盾的。

表 6　单句内不确定性程度较低的矛盾/冲突知识主张识别结果

出版时间	源语句	三元组	矛盾/冲突词
2011.06	20597865. ab. 2 Polymorphism at position 224, C224T, on exon 2 of cathepsin D gene (CTSD) has been associated with an increased risk for Alzheimer's disease (AD) by some investigators, but there have been ***contrary*** findings by others.	Cathepsin D-***predisposes***-Alzheimer's Disease	contrary
2012.04	22543848. ab. 1 In view of the vital role of oxidative stress in the pathogenesis of Alzheimer's disease (AD), the potential of antioxidant supplements to prevent AD have gained much interest, while there are ***conflicting*** results on this topic in recent years.	antioxidants-***prevents***-Alzheimer's Disease; oxidative stress-***causes***-Alzheimer's Disease	conflicting
2013.06	23370341. ab. 1 There were ***conflicting*** results about whether promoter polymorphisms (－2578C/A, －1154G/A) of vascular endothelial growth factor (VEGF) gene is a risk factor of Alzheimer's disease (AD).	VEGF gene-***predisposes***-Alzheimer's Disease	conflicting
2014.06	25114079. ab. 1 Background: Research into Ginkgo biloba has been ongoing for many years, while the benefit and adverse effects of Ginkgo biloba extract EGb761 for cognitive impairment and dementia has been discussed ***controversially***.	EGb 761-***treats***-Dementia	controversially
2015.04	25899425. ab. 1 The role of memantine as a treatment for Alzheimer's disease (AD) has been ***controversial***.	memantine-***treats***-Alzheimer's Disease	controversial
2015.09	26098935. ab. 1 Background: Exposure to aluminum remains a ***controversial*** risk factor for Alzheimer's disease.	aluminum-***predisposes***-Alzheimer's Disease	controversial
2018.07	30045255. ab. 1 Previous studies have indicated that statins use is associated with risk of dementia, but presented ***controversial*** results.	Hydroxymethylglutaryl-CoA reductase inhibitors-***predisposes***-dementia	controversial
2020.01	32021202. ab. 2 The ***consumption*** of alcohol is a controversially discussed risk factor for dementia.	ethanol-***predisposes***-dementia	controversially
2020.12	33395056. ab. 1 Background: Prevention of Alzheimer's disease (AD) with Vitamin D (VD) supplementation has been studied widely, but the results in the literature are very ***conflicting***.	supplementation-***prevents***-Alzheimer's Disease	conflicting

3.3.2　不确定性程度较低的矛盾/冲突知识主张跨语句识别

在去除 3.3.1 中识别出的 125 对三元组后,对剩余的来源于 25 318 条有效语句的 42 650 对三元组进行跨语句识别。根据 2.3 中所设定的规则,共计识别出 87 组表示矛盾/冲突知识主张的三元组对,涵盖 807 条语句。因笔者主要研究阿尔茨海默病(AD)相关的诊断和治疗知识主张,对于普适性的主语(如 diagnosis)或宾语(如 patients、disease、persons),则不在本文的研究关注范围内,需要去除。此外,标题中的语句多是中性的论述(尤其是综述类文章),并不能从中得出结论性的知识主张,故将从标题中抽取得来的部分三元组去除。经过整理,最终得到 64 组与 AD 诊断和治疗相关的表示矛盾/冲突知识主张的三元组对,涵盖 490 条语句。经专家判读,最终确定 36 组、共计 259 对矛盾/冲突三元组。

根据抽取结果,可将与阿尔茨海默病相关的三元组对分为诊断类和治疗类。从表 7 中可以看出,在 AD 的诊断方面,全身麻醉(general anesthesia)、认知受损(impaired cognition)、神经颗粒素(neurogranin)、tau 蛋白(tau proteins)与 AD 的病因、病理或发病机制密切相关,但相关作用机制仍存在较大争议。

国家卫健委 2020 年 12 月份发布的《精神障碍诊疗规范(2020 年版)－阿尔茨海默病》[35]指出,阿尔茨海默病的病理改变以老年斑(Senile Plaques, SP)、神经原纤维缠结(Neurofibrillary Tangle, NFT)和神经元减少为主要特征,而 SP 的中心是 β 淀粉样蛋白(β-amyloid protein, Aβ),Aβ 的生成和清除失衡是神经元变性和痴呆发生的始动因素;NFT 的主要组分是 tau 蛋白,Aβ 会诱导 tau 蛋白过度磷酸化、炎症反应、神经元死亡等

一系列病理过程。同时,神经递质异常(包括乙酰胆碱系统、单胺系统、氨基酸类及神经肽等)也是阿尔茨海默病患者的主要症状。经过比对,本研究识别出的 AD 诊断类相关的矛盾三元组与上述诊疗规范中的内容是基本一致的,说明识别结果较为准确,可以作为后续研究改进的参照着力点。

表 7 跨语句不确定性程度较低的与阿尔茨海默病相关的诊断类矛盾三元组对

序号	主语	谓词(频次)	宾语
1	general anesthesia	predisposes(1) neg_predisposes(1)	Alzheimer's Disease
2	impaired cognition	predisposes(5) neg_predisposes(1)	Alzheimer's Disease
3	neurogranin	predisposes(2) neg_predisposes(1)	Alzheimer's Disease
4	tau proteins	treats(1) predisposes(1)	Alzheimer's Disease

注:谓词后括号中的数字表示该三元组出现的次数

根据《精神障碍诊疗规范(2020 年版)–阿尔茨海默病》中有关 AD 的治疗原则和常用药物内容,可将有关 AD 的治疗方式分为药物治疗和非药物干预两类,药物治疗主要包括:①改善认知的药物,如胆碱酯酶抑制剂、谷氨酸受体拮抗剂(如美金刚);②针对精神行为症状的药物,如抗精神病药、抗抑郁药、心境稳定剂等。非药物干预主要包括环境治疗、感官刺激治疗、行为干预、音乐治疗、舒缓治疗、香氛治疗、认可疗法、认知刺激治疗等多种形式。

与《精神障碍诊疗规范(2020 年版)–阿尔茨海默病》对应,笔者将识别出的用三元组表示的 AD 治疗方式分为药物治疗和非药物干预两类。在药物治疗方面(见表 8),兴奋剂(agonists)、抗糖尿病药(antidiabetics)、非甾体抗炎剂(anti-inflammatory agents, non-steroidal)、抗氧化剂(antioxidants)、阿司匹林(aspirin)、钙通道阻滞剂(calcium channel blockers)、胆碱酯酶抑制剂(cholinesterase inhibitors)、多奈哌齐(donepezil)、羟基甲基戊二酰辅酶 A 还原酶抑制剂(Hydroxymethylglutaryl-CoA reductase inhibitors)、静脉注射免疫球蛋白(immunoglobulins, intravenous)、胰岛素(insulin)、胰岛

表 8 跨语句不确定性程度较低的与阿尔茨海默病相关的药物治疗类矛盾三元组对

序号	主语	谓词(频次)	宾语
1	agonists	treats(3) neg_treats(1)	Alzheimer's Disease
2	agonists	treats(2) neg_treats(1)	impaired cognition
3	agonists	treats(2) neg_treats(1)	schizophrenia
4	antidiabetics	treats(2) neg_treats(1)	Alzheimer's Disease
5	anti-inflammatory agents, non-steroidal	treats(4) neg_treats(2)	Alzheimer's Disease
6	antioxidants	treats(1) predisposes(1)	Alzheimer's Disease
7	aspirin	prevents(1) neg_treats(1)	Alzheimer's Disease
8	calcium channel blockers	treats(1) prevents(1) predisposes(1)	dementia
9	cholinesterase inhibitors	treats(25) neg_treats(1)	Alzheimer's Disease
10	donepezil	stimulates(2) inhibits(2)	acetylcholinesterase
11	Hydroxymethylglutaryl-CoA reductase inhibitors	treats(9) prevents(2) predisposes(1)	Alzheimer's Disease
12	immunoglobulins, intravenous	treats(2) neg_treats(1)	Alzheimer's Disease
13	insulin	augments(1) disrupts(1)	memory
14	insulin-like growth factor I	prevents(1) predisposes(1)	Alzheimer's Disease
15	memantine	treats(1) neg_treats(2)	agitation
16	memantine	treats(23) neg_treats(1)	Alzheimer's Disease
17	memantine	treats(6) neg_treats(1)	dementia
18	memantine	treats(1) neg_treats(1)	lewy body disease
19	monastrol	inhibits(1) neg_inhibits(1)	long-term potentiation
20	pharmaceutical preparations	treats(48) neg_treats(1) predisposes(3)	Alzheimer's Disease
21	pharmaceutical preparations	neg_disrupts(1) disrupts(1)	dementia
22	pharmaceutical preparations	treats(1) neg_treats(1)	mild cognitive disorder
23	proteins	prevents(1) predisposes(1)	Alzheimer's Disease
24	testosterone	treats(1) predisposes(1)	Alzheimer's Disease
25	vitamin D	predisposes(1) neg_predisposes(1)	Alzheimer's Disease
26	vitamin E	treats(2) prevents(3) predisposes(1)	Alzheimer's Disease

郭进京，黄奇. 不确定性环境下的矛盾知识主张识别研究[J]. 图书情报工作,2021,65(20):123－134.

素样生长因子 I(insulin-like growth factor I)、美金刚(memantine)、有丝分裂驱动蛋白 Eg5 抑制剂(monastrol)、蛋白质(proteins)、睾丸素(testosterone)、维他命 D(vitamin D)、维他命 E(vitamin E)等在治疗 AD 或痴呆,改善认知和精神行为症状(如抑郁)的作用方面仍存在矛盾或争议。

在非药物干预方面(见表 9),体育活动(physical activity)、预防性治疗(prophylactic treatment)等方式对 AD、痴呆、认知或精神行为症状的干预治疗作用,仍存在部分矛盾或争议。

表 9 跨语句不确定性程度较低的与阿尔茨海默病相关的非药物治疗类矛盾三元组对

序号	主语	谓词(频次)	宾语	序号	主语	谓词(频次)	宾语
1	physical activity	treats(1) prevents(7) predisposes(2)	Alzheimer's Disease	4	therapeutic procedure	treats(10) neg_treats(4)	dementia
2	prophylactic treatment	treats(1) neg_treats(1)	Alzheimer's Disease	5	therapeutic procedure	treats(5) neg_treats(1)	impaired cognition
3	therapeutic procedure	causes(1) treats(27) neg_treats(1) prevents(1)	Alzheimer's Disease	6	therapeutic procedure	treats(2) neg_treats(1)	indifferent mood

4 讨论

4.1 基于三元组的矛盾/冲突知识主张识别为知识发现提供新方式

目前,学术文献计量研究正经历着从聚焦于学术文献的外部特征到开始关注内容特征,从关注学术文献的句法特征到重视语义特征乃至语用特征的转变[36]。基于引用的前沿识别方法,通过概念共现展示实体间的关系,只体现出相关性,不涉及因果性质。例如,对于药物和疾病的共现,不能从中判断出是指某一药物对疾病有治疗作用还是诱发不良反应事件。基于三元组的知识单元识别,是从文献计量向知识计量的深化,有别于当前使用较为普遍的基于高被引论文的研究热点或研究前沿识别方法,可以体现出实体间是怎样的关系,以及“如何”相互作用,更便于研究人员准确获取实体间的作用机制,有针对性地开展后续研究。识别出目前存在的具有矛盾/争议的知识主张,可为研究人员识别该研究领域尚未达成一致共识的研究前沿提供参考,帮助研究人员解决实际问题,挖掘潜在的新方法、新技术、新知识。同时,在识别矛盾/冲突知识主张的过程中,通过了解每一组中不同知识主张的提出时间(见图 4),可以大致了解某一研究问题随着时间推进而逐步演化的过程,动态了解矛盾/争议知识主张的发展阶段,为解决问题的切入点提供参照。以羟基甲基戊二酰辅酶 A 还原酶抑制剂为例,对其与 AD 的相互作用的研究演变进行展示,如图 4 所示(为便于展示,笔者根据谓词所代表的正向/反向含义及不确定性程度,将“prevents”赋值为 1,“treats”赋值为 2,“predisposes”赋值为－1),2018 年的研究[37]针对他汀类药物羟基甲基戊二酰辅酶 A 还原酶抑制剂与 AD 的治疗或预防作用提出了异议,指出“真菌衍生的他汀类药物(fungus-derived statins)”与 AD 风险增加有关,为深化他汀类药物与 AD 的作用机制提供了不一样的研究结论,有助于推动现有研究的开展。

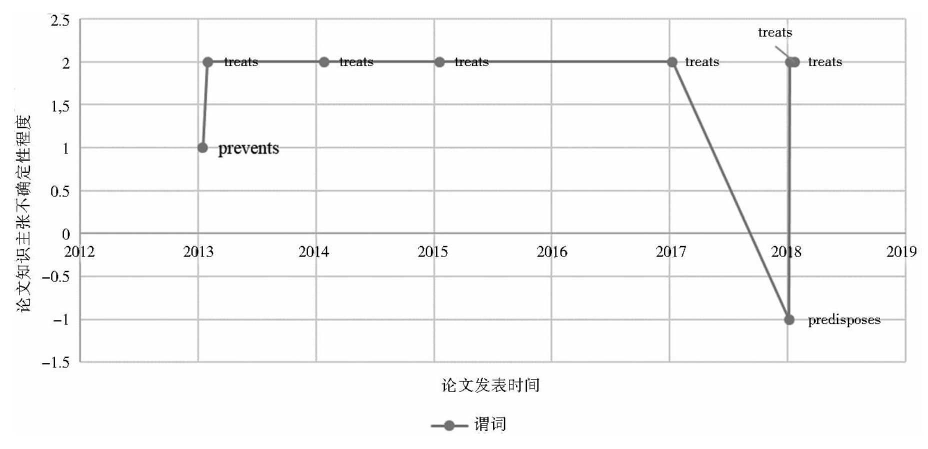


图 4 羟基甲基戊二酰辅酶 A 还原酶抑制剂与 AD 的相互作用研究演变过程

4.2 知识不确定性程度的衡量为知识图谱可信度计算提供参考

本研究以源语句中是否含有"may""could""might"等表示高度不确定性的提示词作为划分依据，将处理后的源语句划分为不确定性程度较高的语句和不确定性程度较低的语句，然后分别进行单句识别和跨语句识别，据此抽取出的具有矛盾/争议的三元组的不确定性程度是不一样的。临床决策过程充满了不确定性，医生的决策过程是从"知识"到"智慧"的过程，解决的是"如何在不确定条件下中做出最佳决策"的问题。笔者在表 7 至表 9 的跨语句识别中，标注出了谓词出现的频次，可以计算某个三元组对在本组知识主张中出现的频率，作为量化特定三元组不确定性程度的依据，也为支撑临床决策的知识图谱的构建提供了确定性程度或置信度参考，同时完整的知识单元可为知识主张提供依据来源。据此，在构建知识图谱时，可以标注出三元组中主语和宾语相互作用的概率，为基于知识图谱的临床决策提供更有说服力的参考。

4.3 矛盾/冲突知识主张的判定离不开专业人员参与

笔者以不确定性程度高低为划分依据，分别进行了矛盾/冲突知识主张的识别，从单句中识别出 15 组知识主张，涉及 18 对三元组；从跨语句中识别出 67 组知识主张，涉及 503 对三元组，以上共计 521 对三元组。这些三元组在经过专家判读后，确定 372 对三元组识别准确，准确率为 71.4%，去掉因识别错误而不能构成矛盾/冲突关系的三元组，最终得到 49 组（涵盖 277 对三元组）矛盾/冲突知识主张。经过分析，该方法仍存在着以下不足：

（1）三元组的识别存在着错误。一类是命名实体识别错误（见表 10），一类是语义关系（即谓词）抽取错误（见表 11）。虽然 SemRep 针对不同类型的语义关系的识别准确率范围为 59% –96%[27]，基本能满足需要，但仍需要专家介入对识别出的三元组进行人工判读和修正。

表 10 实体识别错误举例（以 triacetyluridine 为例）

出版时间	源语句	三元组
2012.02	22086139. ab. 5 Transgenic rodent models over-expressing human -amyloid precursor protein (-APP) and mutant forms of tau have become precious tools to study and understand the pathogenesis of AD at the molecular, cellular and behavioural levels, and to test new therapeutic agents.	triacetyluridine-causes-Alzheimer's Disease

注：三乙酰尿苷（triacetyluridine，规范名称为 2',3',5'-Tri-O-acetyluridine）英文简称 TAU，semrep 错将 tau（实为一种蛋白）映射为了 triacetyluridine，导致实体识别错误

表 11 语义关系提取错误举例（以 Hydroxymethylglutaryl-CoA reductase inhibitors 为例）

出版时间	源语句	三元组
2013.04	23225700. ti. 1 Statins in the prevention of dementia and Alzheimer's disease: a meta-analysis of observational studies and an assessment of confounding.	Hydroxymethylglutaryl-CoA reductase inhibitors-prevents-Alzheimer's Disease

注：该语句是中性论述，并不能从中看出结论性的知识主张，因此在语义关系抽取时会发生错误

（2）表征矛盾/冲突的提示词和识别规则需要优化。尽管相对于已有的研究，笔者扩展了表征矛盾/冲突的提示词的范围（由 3 个扩展至 7 个），但仍待补充和优化。对于跨语句的识别规则，有学者列出了表征矛盾/冲突的谓词对，也有学者将谓词划分为激励类（excitatory）和抑制类（inhibitory）两个相互矛盾/冲突的群组。第一种方法可以清晰的展现识别出的三元组的因果关系，但受限于识别规则的数量，导致识别出的知识主张数量有限。而第二种方法则便于识别出更多的矛盾/冲突知识主张，但不能清晰地获取三元组中的因果关系。这两种方法各有利弊，如何选取应根据研究需要而定。

5 结论

学科领域研究前沿代表着各个学科领域的难点、热点和发展趋势，指引着各国科学发展的方向。2016 年，国务院发布《国家创新驱动发展战略纲要》[38]，指出为实现 2050 年建成世界科技创新强国的宏伟目标，要"面向科学前沿加强原始创新，力争在更多领域引领世界科学研究方向""推进变革性研究，在新思想、新发现、新知识、新原理、新方法上积极进取"。在当前全球许多国家将创新驱动视作谋求竞争优势的核心战略的背景下，识别科学研究前沿、开展变革性研究是提升科学前沿创新能力、赢得发展主动权的必要手段。笔者以阿尔茨海默病为例，根据不确定性原理中的基于矛盾/冲突检测的变革性研究识别方法，制定表征矛

盾/冲突意义的知识主张识别规则,从海量医学文本中识别出以三元组形式表示的、具有矛盾争议的临床知识主张,并与国家卫生健康委员会颁布的《精神障碍诊疗规范(2020 年版)－阿尔茨海默病》中指南内容进行了对比验证,最终结果显示,识别出的矛盾/冲突三元组的主题范围与指南中有关阿尔茨海默病诊断和治疗的内容基本一致,矛盾/知识主张的识别结果与当前研究方向高度契合。该实验是对基于矛盾/冲突的不确定知识主张识别的一次验证,有助于发现与阿尔茨海默病研究相关的尚存争议的潜在研究前沿,可为指导相关研究开展提供参考。未来笔者将进一步向不确定性程度衡量方法优化、从论述句扩展至引证句、丰富表示矛盾/争议的提示词和识别规则等方向扩展该方法。

致谢 感谢盛姝、杨洋、解绮雯、秦新国、尹天恒在研究过程中给予的宝贵意见和建议。

参考文献:

[1] JAMES M. Law reform[J]. London Review, 1835, 2(3): 1－51.
[2] 吴国林. 论知识的客观性[J]. 科学学与科学技术管理, 2000, 21(6):37－39.
[3] LIGHT M, QIU X Y, SRINIVASAN P. The language of bioscience: facts, speculations, and statements in between[C]// HLT-NAACL 2004 workshop: linking biological literature, ontologies and databases. Boston: Association for Computational Linguistics, 2004: 17－24.
[4] The Cell editorial team. What happens next [J]. Cell, 2021, 184(2): 291－292.
[5] Alzheimer's Disease International. Dementia facts & figures[EB/OL]. [2021－02－24]. https://www.alzint.org/about/dementia-facts-figures/.
[6] 中国老年医学会. 倡议书:我国应持续加大对阿尔茨海默病的重视与投入[EB/OL]. [2021－02－24]. https://www.zglnyxxh.com/site/content/3036.html.
[7] KUHN T S, HACKING I. The structure of scientific revolutions: 50th anniversary edition [M]. Chicago: University of Chicago Press, 2012:72－89.
[8] 杜建. 医学知识不确定性测度的进展与展望[J]. 数据分析与知识发现, 2020, 4(10):14－27.
[9] Oxford-Dictionary. Oxford english dictionary online, 2nd edition [EB/OL]. [2021－02－24]. http://www.oed.com/.
[10] CHEN C M, SONG M. Representing scientific knowledge: the role of uncertainty [M]. Switzerland: Springer International Publishing, 2017:1－351.
[11] MCELROY M W . The new knowledge management: complexity, learning, and sustainable innovation[M] Burlington: Elsevier Butterworth-Heinemann, 2003:7.
[12] PETERS K , MARUSTER L, JORNA R J . Knowledge claim evaluation: a fundamental issue for knowledge management[J]. Journal of knowledge management, 2010, 14(2):243－257.
[13] ALAMRI A. The detection of contradictory claims in biomedical abstracts[D]. Sheffield: University of Sheffield, 2016.
[14] MARNEFFE M, RAFFERTY A N, MANNING C D. Finding contradictions in text[C]// Association for Computational Linguistics 2008: Proceedings of the 46th annual meeting of the association for computational linguistics. Stroudsburg: Association for Computational Linguistics, 2008:1039－1047.
[15] DAGAN I, GLICKMAN O, MAGNINI B. The PASCAL recognising textual entailment challenge[C]// Machine learning challenges. evaluating predictive uncertainty, visual object classification, and recognising tectual entailment. Southampton: Springer, 2005: 177－190.
[16] HARABAGIU S M, HICKL A, LACATUSU V F. Negation, contrast and contradiction in text processing[C]// AAAI'06: Proceedings of the 21st national conference on artificial intelligence. Menlo Park: The AAAI Press , 2006: 755－762.
[17] KILICOGLU H, PENG Z, TAFRESHI S, et al. Confirm or refute?: a comparative study on citation sentiment classification in clinical research publications[J]. Journal of biomedical informatics, 2019, 91(3):103123.
[18] MURRAY D, LAMERS W, BOYACK K, et al. Measuring disagreement in science [C]//Proceedings of the 17th international conference of the international society for scientometrics and informetrics(ISSI). Leuven: ISSI Society, 2019: 2370－2375.
[19] ROSEMBLAT G, FISZMAN M, SHIN D, et al. Towards a characterization of apparent contradictions in the biomedical literature using context analysis[J]. Journal of biomedical informatics, 2019, 98(10): 103275.
[20] PINTO J M G, WAWRZINEK J, BALKE W. What drives research efforts? find scientific claims that count! [C]// Proceedings of the 2019 ACM/IEEE joint conference on digital libraries. Washington: IEEE Press, 2019:217－226.
[21] LI X Y, PENG S Y, DU J. Towards medical knowmetrics: representing and computing medical knowledge using semantic predications as the knowledge unit and the uncertainty as the knowledge context[J]. Scientometrics,2021,126(7): 6225－6251.
[22] Turing Center. Open information extraction[EB/OL]. [2021－02－24]. https://openie.allenai.org/.
[23] IHS Markit. Patent solutions for research and discovery[EB/OL]. [2021－02－24]. https://ihsmarkit.com/products/patents-applications.html.
[24] RINDFLESCH T C, FISZMAN M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text[J]. Journal of Biomedical Informatics, 2003, 36(6):462－477.
[25] RINDFLESCH T C, LIBBUS B, HRISTOVSKI D, et al. Semantic relations asserting the etiology of genetic diseases[C]//AMIA Annual symposium proceedings/AMIA Symposium. Rockville: American Medical Informatics Association, 2003:554－558.
[26] FISZMAN M, DEMNER-FUSHMAN D, LANG F M, et al. Inter-

preting comparative constructions in biomedical text[C]//Biological, translational, and clinical language processing. Prague: Association for Computational Linguistics, 2007:137 – 144.
[27] KILICOGLU H , ROSEMBLAT G , FISZMAN M, et al. Broad-coverage biomedical relation extraction with SemRep[J]. BMC Bioinformatics, 2020, 21(1):1 – 28.
[28] KHAN M A, YUAN Y, IQBAL U, et al. No association linking short-term proton pump inhibitor use to dementia: systematic review and meta-analysis of observational studies[J]. American journal of gastroenterology, 2020, 115(5):671 – 678.
[29] MADHAV D, VENKAT N, SACHIN S, et al. Proton pump inhibitors do not increase the risk of dementia: a systematic review and meta-analysis of prospective studies[J]. Diseases of the esophagus, 2020,33(10):1 – 9.
[30] HAYNES R B. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions[J]. Evidence based medicine, 2006, 11(6):162 – 164.
[31] SMALL H. What makes some scientific findings more certain than others? a study of citing sentences for low-hedged papers[C] //Proceedings of the 17th international conference on scientometrics and informetrics(ISSI). Leuven: ISSI Society, 2019:554 – 560.
[32] LUCCHETTA R C, MATA B P M D, MASTROIANNI P D C. Association between development of dementia and use of benzodiazepines: a systematic review and meta-analysis[J]. Pharmacotherapy: the journal of human pharmacology and drug therapy, 2018, 38(10):1010 – 1020.
[33] SU C, UNDERWOOD W, RYBALCHENKO N, et al. ERK1/2 and ERK5 have distinct roles in the regulation of brain-derived neurotrophic factor expression[J]. Journal of neuroscience research, 2011, 89(10):1542 – 1550.
[34] XIE B, ZHOU H M, LIU W X, et al. Evaluation of the diagnostic value of peripheral BDNF levels for Alzheimer's disease and mild cognitive impairment: results of a meta-analysis[J]. International journal of neuroscience, 2019, 130(1):1 – 14.
[35] 国家卫生健康委办公厅. 阿尔茨海默病的诊疗规范(2020 年版)[J]. 全科医学临床与教育, 2021, 19(1):4 – 6.
[36] 卢超,章成志,王玉琢,等. 语义特征分析的深化——学术文献的全文计量分析研究综述[J]. 中国图书馆学报, 2021, 47(2):110 – 131.
[37] SINYAVSKAYA L, GAUTHIER S, RENOUX C, et al. Comparative effect of statins on the risk of incident Alzheimer disease[J]. Neurology: official journal of the American academy of neurology, 2018, 90(3):179 – 187.
[38] 中共中央国务院印发《国家创新驱动发展战略纲要》[EB/OL]. [2021 – 02 – 24]. http://www.gov.cn/xinwen/2016-05/19/content_5074812.htm.

作者贡献说明:
郭进京:论文选题、数据处理与初稿撰写;
黄奇:论文结构优化,论文修改。

Research on Identification of Contradictory Knowledge Claims Under the Uncertainty Environment

Guo Jinjing Huang Qi
School of Information Management, Nanjing University, Nanjing 210023

Abstract: [Purpose/significance] Scientific research, especially medical research, is full of uncertainty. Identifying contradictory knowledge claims in research will help identify "scientific divergence/inconsistent scientific conclusions", promote identification of potentially transformative scientific research and improve related research. [Method/process] Taking Alzheimer's disease as an example, this article used SemRep to extract triples from the PubMed abstract data. Based on the knowledge claim identification rules that represented contradictory meanings, this article classified source sentences according to the degree of uncertainty and used single-sentence recognition and cross-sentence recognition respectively to identify contradictory medical research knowledge claims expressed in the form of triples. [Result/conclusion] A total of 49 groups (involving 277 pairs of triples) contradictory knowledge claims are identified from 6574 PubMed medical abstracts . There are still some controversies and contradictions in the diagnosis and treatment of Alzheimer's disease, which need to be further verified. The identification of contradictory knowledge claims provides new ideas for the frontier discoveries of potentially transformative medical research, which can be used for knowledge discovery based on knowledge measurement and provide references for the reliability calculation of knowledge graph.

Keywords: knowledge discovery uncertainty contradictory knowledge SPO Alzheimer's disease